Regularities in football goal distributions

L. C. Malacarne & R. S. Mendes
Departamento de Física, Universidade Estadual de Maringá, Avenida Colombo 5790, 87020-900, Maringá-PR, Brazil

Besides of complexities concerning to football championships, it is identified some regularities in them. These regularities refer to goal distributions by goal-players and by games. In particular, the goal distribution by goal-players it well adjusted by the Zipf-Mandelbrot law, suggesting a conection with an anomalous decay.

Regularity in some complex systems can sometimes be identified and expressed in terms of simple laws. Typical examples of such situations are found in a wide range of contexts as the frequency of words in a long text [1], the population distribution in big cities [2–4], forest fires [5], the distribution of species lifetimes for North American breeding bird populations [6], scientific citations [7,8], www surfing [9], ecology [10], solar flares [11], economic index [12], epidemics in isolated populations [13], among others. Here, universal behaviours in the most popular sport, the football, are discussed. More precisely, this work focuses on regularities in goal distribution by goal-players and by games in championships. Furthermore, the goal distribution by goal-players is connected with an anomalous decay related to the Zipf-Mandelbrot [1,14] law and with Tsallis nonextensive statistical mechanics [15–18].

In many contexts, it is common that few phenomena with high intensity arise, and so do many ones with low intensity. For instance, a long text generally contains many words that are employed in few opportunities and a small number that occurs largely [1]. The above mentioned systems are good examples too. In particular, this kind of behaviour usually occurs in football championships, because there are many players that make few goals in contrast with the topscorers.

A detailed visualization of this behaviour can be well illustrated by considering some of the most competitive and traditional championships of the world. Our particular choice of championships has been done guided by the criterion of easy accessibility of the corresponding data to anyone [19,20]. Therefore, we consider, here, some of the main league football championships from Italy, England, Spain and Brazil [21]. Each of these championships has the participation of about twenty teams, contains around three hundred games, and approximately eight hundred goals [22]. In Fig. 1 we exhibit data of these championships. In these graphics, the abscissa presents the number of goals x divided by an average of goals m (total number of goals per total number of goal-players), and the ordinate indicates the quantity $N(x)$ of players with x goals divided by the quantity of players with one goal, $N(1)$. The regular shape of the graphics presented in Fig. 1 suggests a general law to describe the distribution of goals.

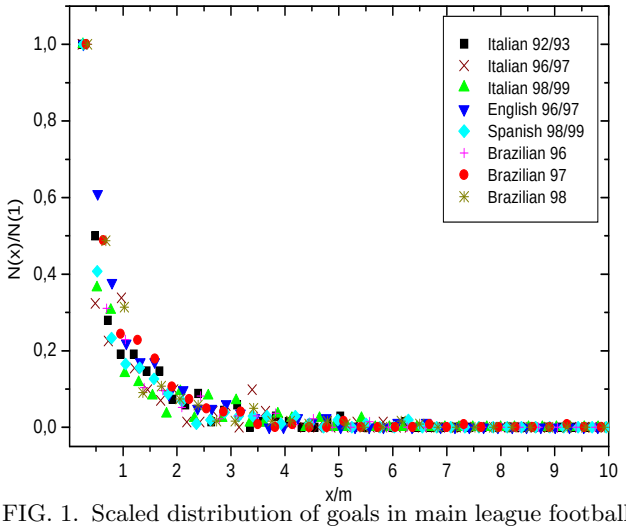

FIG. 1. Scaled distribution of goals in main league football championships from Italy, England, Spain and Brazil. The ordinate $N(x)$ is the number of players with x goals divided by the number of players with one goal, $N(1)$, and the abscissa is the number of goals divided by the average of goals m. These scaled data indicate a regularity in the goal distribution for goal-players. The European championships start in one year and finish in the subsequent year, and the Brazilian championships start and finish in the same year.

In the study of the majority of the previously cited systems, the Zipf's law [1], $N(x) = a/x^b$, arises naturally, at least in part of the analysis. In the Zipf's law, a and b are constants and x is the independent variable. In order to give a better adjustment to a large part of the data, and based on information theory, Mandelbrot [14] proposed $N(x) = a/(c+x)^b$ as a generalization of the Zipf's law, with a, b, and c all being constants. This Zipf-Mandelbrot's distribution also arises in the context of a generalized statistical mechanics proposed some years ago [15,16,18], equivalently rewritten as

$$N(x) = N_0[1-(1-q)\lambda x]^{\frac{1}{1-q}} , \qquad (1)$$

where N_0, λ, a q are real parameters. In addition, this function satisfies an anomalous decay equation,

$$\frac{d}{dx}\left(\frac{N(x)}{N_0}\right) = -\lambda\left(\frac{N(x)}{N_0}\right)^q . \qquad (2)$$

The parameter q can be considered as a measure of how anomalous the decay is. In particular, equation (1) is reduced to the usual exponential decay, $N(x) = N_0 \exp(-\lambda x)$, in the limit $q \to 1$.

Motivated by these physical connections, we employ the distribution (1) to adjust the goals data. Following the construction of Fig. 1, we use the number of goalplayers with one goal, $N(1)$, and the average goal number by goalplayer, m, to eliminate N_0 and λ. Furthermore, it is a good approximation to replace the discrete average with a continuous one in the present analysis, *i.e.*,

$$m = \frac{\int_0^\infty xN(x)}{\int_0^\infty N(x)} = \frac{1}{\lambda(3-2q)} \qquad (q < 3/2). \qquad (3)$$

Thus, the distribution of goals dictated by equation (1) can be rewritten as

$$N(x) = N(1)\frac{\left[1-\frac{(1-q)}{(3-2q)m}\right]^{\frac{1}{q-1}}}{\left[1-\frac{(1-q)}{(3-2q)m}x\right]^{\frac{1}{q-1}}}, \qquad (4)$$

where q becomes the unique parameter that remains to be adjusted, since $N(1)$ and m are obtained directly from the data. Fig. 2 illustrates applications of equation (4) for four championships, indicating therefore the goodness of the formula (4). The same conclusion is obtained in the other championships showed in Fig. 1. Here, $q = 1.33$ was employed as approximated value, leading to the Zipf-Mandelbrot's exponent $b \approx 3$. In this way, $q \approx 1.33$ can be interpreted as the universal parameter for this kind of championships. Also, it is interesting to remark that $b \approx 3$ occurs in the distribution of scientific citations [7,8].

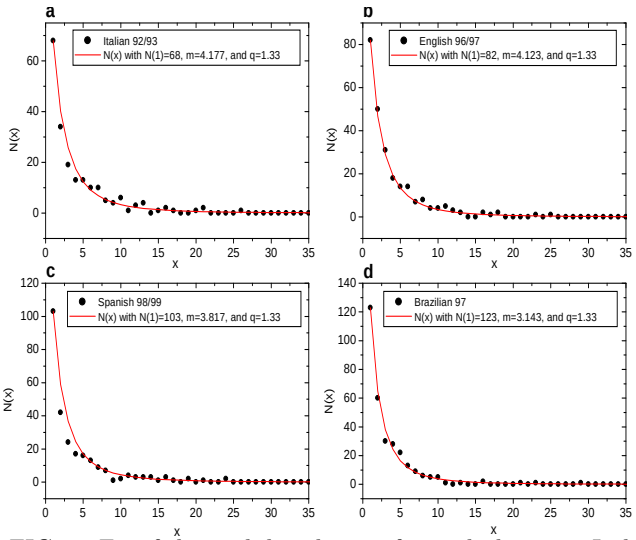

FIG. 2. Fit of the goal distribution for goal-players in Italian (**a**), English (**b**), Spanish (**c**), and Brazilian (**d**) championships. The black circles are the goal data and the solid line is the fitting with the $N(x)$ distribution being given in equation (4) with $q = 1.33$.

Other kinds of regularities in goal distributions can be identified, but with different behaviours. This fact can be verified in the distribution of goals per game. Proceeding in a similar way as done in Fig. 1, it is considered normalized scale distribution of games and goals. In this case, the abscissa is the number x of goal divided by M, the mean goal per game of a championship (the number of goals of a championship divided by the corresponding number of games). In addition, the ordinate is given by the number of games with x goals of a championship divided by the number of games of the corresponding championship. Fig. 3 contains this kind of graphics illustrating this regular behaviour by considering, again, the main league football championships from Italy, England, Spain and Brazil. As one can see, this figure strongly suggests a regularity in the distribution of goals for distinct championships.

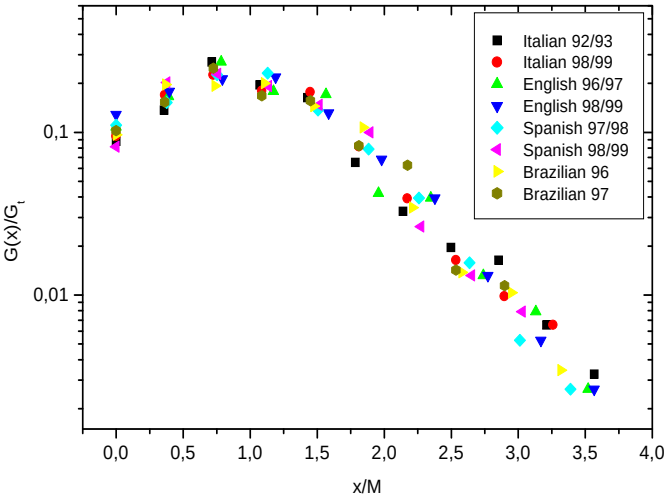

FIG. 3. Semi-log graphic of scaled distributions of goals per game is illustrated by considering championships from Italy, England, Spain and Brazil. For each championship, x is the number of goals per game, M is the average of goals per game, $G(x)$ is the number of games with x goals, and G_t is the total number of games.

Besides the numberless factors, including fluctuation due to relatively small number of teams and games that are present in a football championship, regularities arise in the goal distributions. In particular, the goal distribution for players that make goals are well adjusted by a Zipf-Mandelbrot law, suggesting a connection with ubiquitous phenomena such as anomalous diffusion.

[1] G. K. Zipf, *Human Behavior and the Principle of Least Effort* (Addison-Wesley, Cambridge, MA, 1949).

[2] A. H. Makse, S. Havlin, H. E. Stanley, *Nature* **377**, 608 (1995).
[3] D. H. Zanette, S. C. Manrubia, *Phys. Rev. Lett.* **79**, 523 (1997).
[4] M. Marsilli, Y.-C. Zhang, *Phys. Rev. Lett.* **80**, 2741 (1998).
[5] B. D. Malamud, G. Morein, D. L. Turcotte, *Science* **281**, 1840 (1998).
[6] T. H. Keit, H. E. Stanley, *Nature* **393**, 257 (1998).
[7] S. Redner, *Eur. Phys. J. B* **4**, 131 (1998).
[8] C. Tsallis, M. P. Albuquerque, *To be published in Eur. Phys. J. B (cond-mat/9903433).*
[9] B. A. Huberman, P. L. T. Pirolli, J. E. Pitkow, R. M. Lukose, *Science* **280**, 95 (1998).
[10] J. R. Banavar, J. L. Green, J. Harte, and A. Maritan, Phys. Rev. Lett. **83**, 4212 (1999).
[11] G. Boffetta, V. Carbone, P. Giuliani, P. Veltri, and A. Vulpiani, Phys. Rev. Lett. **83**, 4662 (1999).
[12] R. N. Mantegna, H. E. Stanley, *Nature* **376**, 46 (1995).
[13] C. J. Rhodes, M. Anderson, *Nature* **381**, 600 (1996).
[14] B. B. Mandelbrot, *The Fractal Geometry of Nature* (Freeman, New York, 1977).
[15] C. Tsallis, *J. Stat. Phys.* **52**, 479 (1988).
[16] E. M. F. Curado, C. Tsallis, *J. Phys. A: Math. Gen.* **24**, L69 (1991); Corrigenda: **24**, 3187 (1991) and **25**, 1019 (1992).
[17] S. Denisov, *Phys. Lett. A* **235**, 447(1997).
[18] C. Tsallis, R. S. Mendes, A. R. Plastino, *Physica A* **261**, 534 (1998).
[19] http://www.risc.uni-linz.ac.at/non-official/rsssf/ .
[20] http://www.globosat.com.br/sportv/ .
[21] Since the complete set of data for all championships of these countries is not easely avaliable, we have used data that are both rencent and easely accessible.
[22] In the European championships, the leagues are organized in such way that each team plays two games with each of the other teams. In the Brazilian championships, each team plays only once with each one of the others. After this, the eight best teams go to the playoff. Nevertheless, the number of games of the Brazilian championships in the playoff is much smaller than in the classification phase, and it doesn't affect the analysis when compared with European ones.

ACKNOWLEDGMENTS

One of us, Mendes, R. S., thanks partial finantial support by CNPq (Brazilian Agency).